Journal of
Applied Remote Sensing

Simple method for retrieving leaf area index from Landsat using MODIS leaf area index products as reference

Feng Gao
Martha C. Anderson
William P. Kustas
Yujie Wang

Simple method for retrieving leaf area index from Landsat using MODIS leaf area index products as reference

Feng Gao,[a] Martha C. Anderson,[a] William P. Kustas,[a] and Yujie Wang[b]
[a]U. S. Department of Agriculture, Agricultural Research Service,
Hydrology and Remote Sensing Laboratory, Beltsville, Maryland 20705
Feng.Gao@ars.usda.gov
[b]University of Maryland Baltimore County, Goddard Earth Sciences and Technology Center,
Baltimore, Maryland 21228

Abstract. Leaf area index (LAI) is a key parameter in most land surface models. Models that operate at multiple spatial scales may require consistent LAI inputs at different spatial resolutions or from different sensors. For example, the atmosphere-land exchange inverse model and associated disaggregation algorithm (DisALEXI) use the moderate resolution imaging spectroradiometer (MODIS) LAI product to model fluxes at regional scales (1- to 10-km grid resolution), and Landsat-based LAI to disaggregate to field scale (30-m grid). In order to make a MODIS-consistent LAI product from Landsat imagery for this combined scheme, a simple reference-based regression tree approach was developed. This approach uses homogeneous and high-quality LAI retrievals from MODIS as references to develop a regression tree relating these MODIS LAI samples to Landsat surface reflectances. Results show that the approach can produce accurate estimates of LAI from Landsat, as evaluated using field measurements collected during the soil moisture experiment of 2002, conducted in central Iowa during a period of rapid vegetation growth. The coefficient of determination (r^2) computed between Landsat retrievals and field measurements was 0.94 at the field scale, with an overall mean bias error (MBE) of −0.07 and mean absolute difference (MAD) of 0.23. MAD values of 0.17 and 0.32 were obtained for low to moderate LAI (0–3) and high LAI (>3), respectively, with some underestimation for the high LAI (MBE = −0.28). The LAI maps retrieved from Landsat were consistent with the MODIS estimates when aggregated to coarser scales. MAD computed between Landsat- and MODIS-derived LAI ranged from 0.07 to 0.83 for different Landsat dates, with no significant bias compared to MODIS high-quality retrievals. This approach demonstrates a simple framework for producing MODIS-consistent LAI from Landsat data for modeling the land surface at different spatial scales.
[DOI: 10.1117/1.JRS.6.063554]

Keywords: leaf area index; Landsat; MODIS; SMEX02 field campaign.

Paper 12013 received Jan. 13, 2012; revised manuscript received Jun. 2, 2012; accepted for publication Jun. 14, 2012; published online Jul. 18, 2012.

1 Introduction

Leaf area index (LAI) is a key biophysical parameter in most land-surface models, governing partitioning of energy, carbon, and water fluxes between the soil and canopy components of the land-surface system.[1] Model fluxes and surface behavior can be highly sensitive to LAI inputs, necessitating development of accurate methods for high-quality LAI retrieval, generally using remote sensing data.[2–5] For global to continental scale applications, coarse-resolution (here defined as kilometer-scale) data, such as from the moderate resolution imaging spectroradiometer (MODIS), provide sufficient spatial detail on a regular basis.[6] MODIS is a key instrument aboard NASA's Terra (morning) and Aqua (afternoon) satellites and acquires data globally twice per day. The MODIS LAI product has been validated and analyzed with independent field

0091-3286/2012/$25.00 © 2012 SPIE

measurements.[7–13] For applications at regional to local scales, however, often finer-resolution LAI data are required, such as maps resolving individual agricultural fields.

For example, Anderson et al.[14,15] describe a multi-scale surface energy balance system that maps evapotranspiration (ET) and other surface fluxes from field to global scales, using land surface temperature and vegetation cover information from a suite of satellite sensors. At the continental to global scales, the atmosphere-land exchange inverse (ALEXI) model uses time-differential measurements of surface temperature from geostationary satellites along with MODIS LAI to generate flux maps at resolutions of 3 to 10 km.[14,16] For more detailed spatial analyses, such as mapping variability in water use across a watershed or between individual farm fields, an ALEXI disaggregation approach (DisALEXI) can be applied using temperature and cover information from polar orbiting sensors (i.e., Landsat or MODIS) to map fluxes at 30 to 1000 m resolution.[4,17] Experiments in fusing Landsat- and MODIS-derived flux fields may lead to methods for generating reliable ET information at both high spatial and temporal resolutions.[15,18]

For these applications, it is beneficial that the Landsat-derived LAI be consistent with the MODIS LAI fields at the 1-km scale; that is, that the aggregated LAI from the fine-resolution sensor should agree with the coarse-resolution LAI. This maximizes self-consistency in retrieved fluxes between the different spatial resolutions. However, standard LAI retrieval approaches (either empirical or physical) for Landsat data will likely produce LAI fields that are inconsistent with the MODIS LAI product. Even using the same MODIS LAI retrieval algorithm, LAI could be different for Landsat and MODIS due to the differences in spectral response functions, land-cover classifications, and required tuning of parameters in preprocessing procedures.[5] These inconsistencies in LAI inputs will cause discrepancies in the flux fields produced at different spatial resolutions, particularly when there exists significant sub-pixel variation in cover fraction and moisture,[19] and will degrade the potential for multi-sensor data fusion techniques.

This paper describes a simple reference-based approach to retrieving LAI from Landsat reflectance data using the MODIS LAI product as a reference. Our objective is to create a MODIS-consistent LAI product from Landsat data for use in multi-scale modeling systems such as ALEXI/DisALEXI. We start by describing the algorithm and experiment site, and then present results, followed by conclusions and a brief discussion of plans for future method development.

2 Methodology

The reference-based methodology described here is an empirical approach that uses high-quality retrievals from the MODIS LAI product as a reference for retrieving LAI from Landsat reflectance maps. Figure 1 is an illustration of the processing framework for the reference-based approach. Similar to the approach proposed to build consistent surface reflectance from multiple medium-resolution sensors,[20] this methodology uses the coarse-resolution MODIS LAI product as a reference to produce consistent Landsat-resolution LAI. In this processing framework,

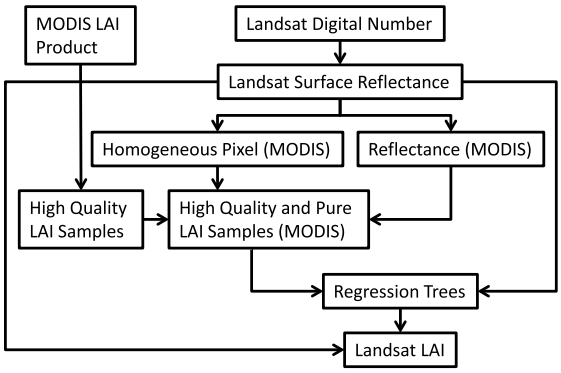


Fig. 1 The processing diagram of the reference-based approach, which uses high-quality MODIS LAI retrievals as references to retrieve a consistent LAI product from Landsat data.

Landsat data in digital number form are first calibrated and atmospherically corrected to surface reflectances using the Landsat ecosystem disturbance adaptive processing system (LEDAPS).[21] Landsat surface reflectances are aggregated to match coarse-resolution MODIS data. Coarse-resolution LAI pixels are extracted from the MODIS LAI product over the Landsat scene. High-quality MODIS LAI samples retrieved from the main algorithm are selected based on the product quality flags. The selected MODIS LAI samples are further screened to exclude pixels with high sub-pixel variability in Landsat reflectance, thus ensuring that only homogeneous MODIS pixels are used in the training process. As LAI has a nonlinear relation to spectral reflectance and VI, a multi-variant regression tree approach [cubist by RuleQuest (The mention of trade names of commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.)] is used in the processing. LAI and surface reflectance samples at MODIS pixel resolution are used to train the Landsat resolution model. The derived regression trees are finally applied to the Landsat surface reflectances (Landsat resolution) to produce Landsat-scale LAI maps.

Empirical LAI retrieval approaches relate LAI to arithmetic combinations (indices) of spectral bands. These indices may be computed from visible, near, or middle infrared bands.[22–24] Instead of using the derived indices, in this experiment the surface reflectances of Landsat 5 Thematic Mapper (TM) or Landsat 7 Enhanced Thematic Mapper Plus (ETM+) bands 2 to 5 and 7 were used directly to build the regression trees. Landsat band 1 was excluded in building the regression tree due to its low signal-to-noise ratio after atmospheric correction.[21] However, all Landsat bands (1 to 5 and 7) were used in determining homogeneous MODIS samples.

Similar to the empirical approaches, the accuracy of the regression tree approach relies on the quality and distribution of samples. To include a wider range of sampled LAI, data from multiple MODIS and Landsat image pairs spanning a period of rapid vegetation growth were combined to build a regression tree and the same tree was then applied to all Landsat scenes used in this paper.

In this processing framework, several factors that impact LAI retrievals need to be considered.

2.1 *MODIS High-Quality LAI Retrievals*

The MODIS LAI product provides quality control flags for each pixel. In the MODIS LAI algorithm, LAI can be retrieved either from the physically based radiative transfer model (main algorithm) or the empirical-based (vegetation index) approach (backup algorithm).[3] A summary of quality analysis and validation activities associated with the MODIS LAI product indicates that MODIS LAI retrieved from the radiative transfer model with the best quality flags can reach an accuracy of 0.3 LAI for cropland[8,10] and 0.5 LAI for needleleaf forest.[7,12] The overestimation of LAI identified in early MODIS LAI collections has been addressed in Collection 4 processing and was further refined in Collection 5 processing.[9,11] Generally, the quality-control flags embedded in the MODIS LAI product reflect the retrieval quality reasonably well.[7–12] The MODIS LAI from the main algorithm shows realistic spatial variation at the continental scale.[13] To ensure that only the best quality of data are being used in the regression tree training, LAI retrievals were only selected that were generated with the best quality radiative transfer model (main algorithm as flagged in MODIS LAI product in SCF_QC bit). As LAI retrievals normally are less reliable when the spectral signals tend to saturate at moderately high values of LAI, retrievals with a saturation quality flag (SCFQC = 1) were excluded from sampling.

2.2 *Pure Homogeneous MODIS Pixels*

LAI has a nonlinear relation to spectral reflectance and derived vegetation indices (VI).[3,4] Suppose this relationship can be expressed as

$$\mathrm{LAI} = f(S), \tag{1}$$

where LAI represents leaf area index, S represents surface reflectance for a suite of spectral bands, and f is a nonlinear function that relates LAI and surface reflectance S. The function f could represent a simple empirical model or a complex physical model.

In general there are two ways to compute LAI at MODIS resolution using Landsat data. One is to compute LAI from Landsat data and then linearly aggregate these fine-scale LAI values to MODIS resolution as:

$$\mathrm{LAI}_M = \frac{\sum_{i=1}^{n} \mathrm{LAI}_L(i)}{n} = \frac{\sum_{i=1}^{n} f_L[S_L(i)]}{n}, \tag{2}$$

where subscripts M and L represent MODIS and Landsat, respectively, i is the index associated with each Landsat pixel within a given MODIS pixel cell, and n is the total number of Landsat pixels in this MODIS pixel cell.

Another way to retrieve LAI at MODIS resolution is to first aggregate the Landsat surface reflectances to MODIS resolution and then compute LAI at that scale:

$$\mathrm{LAI}_M = f_M(S_M) = \frac{f_M[\sum_{i=1}^{n} S_L(i)]}{n}. \tag{3}$$

The MODIS LAI values from these two approaches should be equal, which leads to

$$f_M\left[\sum_{i=1}^{n} S_L(i)\right] = \sum_{i=1}^{n} f_L[S_L(i)]. \tag{4}$$

As f is a nonlinear function, the function composition is not commutative here and thus we can conclude $f_M \neq f_L$. This inequivalence of LAI functions at different spatial resolutions has been noted in early studies.[4,25,26] However, there are two situations that can make them equivalent. One is the situation in which the nonlinear function f can be segmented as multiple linear functions and all Landsat surface reflectance $S_L(i)$ happen to fall along a similar linear segment. Within the small segment, the function f can be treated as the same for both Landsat and MODIS. This requires that all the Landsat surface reflectance $S_L(i)$ in a MODIS cell are very similar. The other situation is that in which the Landsat surface reflectances $S_L(i)$ in the MODIS pixel cell are identical (homogeneous) so that the nonlinear function derived at MODIS resolution can be used at the Landsat resolution directly. In either situation, we require that MODIS LAI samples have low sub-pixel variation in surface reflectances from Landsat (or pure MODIS pixels, as determined at the Landsat scale). The sub-pixel variance can be computed from Landsat surface reflectances inside each MODIS pixel cell. In this study, the coefficients of variation (CV, ratio of the standard deviation to the mean value) for the MODIS pixels were computed and averaged among all spectral bands. The CV describes the relative variation of a MODIS pixel in a Landsat resolution. The smaller the CV is, the purer the MODIS pixel will be. If the band-averaged CV of a MODIS pixel is less than a threshold, the MODIS pixel is considered a homogeneous sample. The thresholds of the CV (indicted below) of surface reflectances should be adjusted to maintain the purity of the MODIS pixels while ensuring there are enough MODIS samples to effectively train the regression tree.

2.3 *Regression Tree for LAI*

Many empirical approaches retrieve LAI using spectral VI.[2,3,22–24] Although the approaches are simple to implement, neglecting vegetation structure and other effects related to cover type is likely to cause a large uncertainty in LAI mapping when extending the algorithm beyond the calibration site. This uncertainty can be reduced by considering cover types in the LAI–VI relations.[3,27] As the cover type information comes from the classification map that is produced from the spectral reflectance signals, it is reasonable to use reflectance from multiple bands in the LAI retrieval. In this study, we rely on a multi-variant regression tree approach (cubist from RuleQuest) to distinguish cover types and build LAI-reflectance relations based on the intrinsic data features. The cubist regression tree method is a data mining approach that builds rule-based predictive multivariate linear regression models based on available samples.[28] The model

tree is constructed using divide-and-conquer method recursively based on entropy measures. The cubist (or M5 in early open version) algorithm combines regression trees and classification to make model trees.[28] The regression tree approach has been widely used in different research fields as well as in remote sensing, such as in the construction of the national land cover database (NCLD)[29] and the MODIS continuous field product.[30]

The LAI samples can be weighted according to their importance or uncertainty. In this study, the weight of each sample was defined as the inverse of the CV of surface reflectances within the MODIS pixel. To balance the samples from different dates, the weights of samples were normalized using Eq. (5) such that the total weights (or contributions) from each date are equal.

$$w_{ij}{}' = \frac{w_{ij}}{\sum w_{ij}} \cdot T_w, \tag{5}$$

where w_{ij} is the adjusted weight for sample i at date j, w_{ij} is the original weight, and T_w is the defined total weight (100 in this paper) for each date. The original weight w_{ij} is computed by

$$w_{ij} = \frac{1}{CV_{ij}} = \frac{\mu_{ij}}{\sigma_{ij}} \tag{6}$$

where μ_{ij} and σ_{ij} are the mean and standard deviation of Landsat surface reflectances within a MODIS pixel.

The number of rules used in cubist training was tested for the study area. It was found that the average prediction error (mean absolute difference) and correlation coefficient from training samples became stable when the number of rules reached 5. In the process of constructing the regression tree (classification), Landsat band 3 (red) and band 4 [near infrared (NIR)] were the only two bands used, which indicates the importance of red and NIR bands in LAI retrievals. In the process of building multivariate linear regression, contributions from each band differed. Landsat band 3 and band 4 were used in all linear regressions (100%), while bands 7, 5, and 2 were only used in 72%, 67%, and 56% of the regressions, respectively. The following two example rules (tree leaves) from the established cubist regression tree in this paper show conditions and linear regressions for two conditions. Rule 1 contains 8882 samples with an average LAI of 0.41. This rule applies to samples of sparse vegetation or bare soil, which is defined by band 4 (NIR) reflectance <0.2621. The LAI is computed from linear regression based on bands 2, 3, and 4 reflectance values. Rule 5 contains 1624 samples with mean LAI of 3.55. This rule applies to dense vegetation and is defined by both band 3 (<0.04553) and band 4 (>0.43903). LAI is computed from bands 2, 3, 4, and 7.

```
Rule 1: [8882 cases, mean 0.41, range 0.1 to 1.7, est err 0.06]
   if
      B4 ≤ 0.26212
   then
      LAI = 0.02 + 6.2B4 – 4.9B3 – 4.2B2

Rule 5: [1624 cases, mean 3.55, range 0.8 to 4.7, est err 0.27]
   if
      B3 ≤ 0.04553
      B4 ≤ 0.43903
   then
      LAI = 1.14 – 39.8B3 + 12.2B4 – 34.6B2 – 5.9B7
```

3 Study Site and Data Processing

This simple reference-based LAI retrieval methodology was tested using satellite and ground data collected during the soil moisture experiment of 2002 (SMEX02) field campaign, focused on the period from June 15 to July 8. Within the Walnut Creek Watershed in central Iowa, 21 corn and 10 soybean fields were selected as sites for intensive soil moisture and vegetation sampling.[4] LAI was measured in four sampling rounds (June 15–19, June 27–30, July 2–3, and July 5–8) using LAI-2000 (LI-COR) Plant Canopy Analyzers at 12 field sites. The remaining 19 sites were sampled twice, in rounds 2 and 4. The full sampling interval covered the period from emergence to tasseling/full bloom in many corn/soybean fields. Average field size was approximately

Table 1 Available Landsat scenes and MODIS LAI products over the SMEX02 study area in 2002 (TM: Landsat-5 Thematic Mapper; ETM+: Landsat-7 Enhanced Thematic Mapper Plus).

Data pair	Landsat	Sensors	WRS-2	Landsat quality	MODIS LAI
1	134 (5/14)	ETM+	p26r31	clear	129–136 (5/9–5/16)
2	158 (6/7)	TM	p26r31	clear	153–160 (6/2–6/9)
3	174 (6/23)	TM	p26r31	clear	169–176 (6/18–6/25)
4	182 (7/1)	ETM+	p26r31	clear	177–184 (6/26–7/3)
5	189 (7/8)	ETM+	p27r31	10% cloud	185–192 (7/4–7/11)
6	198 (7/17)	ETM+	p26r31	20% cloud	193–200 (7/12–7/19)
7	214 (8/2)	ETM+	p26r31	clear	209–216 (7/28–8/4)
8	238 (8/26)	TM	p26r31	40% cloud	233–240(8/21–8/28)
9	245 (9/2)	TM	p27r31	clear	241–248 (8/29–9/5)

500×500 m. For each of the 21 fields, vegetation sampling was conducted with multiple repetitions within three to four sampling sites with a characteristic observation scale of 10×10 m. Anderson et al.[4] compared the field LAI observations with LAI retrievals based on VI maps acquired with Landsat and an airborne imaging system, suggesting a typical retrieval error of about 0.6 LAI.[4]

The SMEX02 field sites were located within two overlapping Landsat worldwide reference system (WRS)-2 scenes (paths 26–27 and row 31), thus increasing the available Landsat scenes for use in this study. Nine clear Landsat-5 and Landsat-7 scenes are available from May 14 to September 2, 2002. Table 1 lists the selected Landsat dates along with coincidental periods for the MODIS Terra eight-day composite LAI product (MOD15A2). Note that the combined Terra and Aqua four-day composite LAI product (MCD15A3) is only available after day of year (DOY) 185 (July 5–8) and thus was not used in sampling and training. The four-day composite LAI was only used and evaluated to compare with the standard eight-day product.

Landsat data were calibrated and atmospherically corrected to surface reflectances using the LEDAPS approach. All images listed in Table 1 are clear over the SMEX02 measurement sites (a small subset of a Landsat scene). The Landsat ETM+ image on July 17 shows about 20% cloud contamination for the top and bottom portions of the scene, but not over the measurement sites. Landsat ETM+ on July 8 is an adjacent path (path 27 and row 31) to the primary scene (path 26 and row 31), and has about 10% cloud contamination in the top portion of the scene, which is not over the SMEX02 study area. The Landsat TM scene on August 26 has about 40% cloud contamination, but again not over the SMEX02 domain. Since automatic detection of clouds and cloud shadows for Landsat imagery is still challenging, the cloudy Landsat scenes were excluded from the sampling and training process to maintain high image quality. However, LAI retrievals were applied to all of these scenes using the combined regression tree built from the other six image pairs.

The *in situ* LAI measurements from the four sampling rounds were linearly interpolated or extrapolated to three Landsat dates on June 23, July 1, and July 8 using the closest sampling events. The temporally interpolated data were used for direct comparison between Landsat retrievals (30 m) and site measurements at the observational scale (10 m). Estimates of average field-scale LAI were aggregated from empirical VI-based LAI retrievals using aircraft imagery and Landsat data in previous work.[4] The field scale ground observations on the four nominal measurement dates (June 16, June 23, July 1, and July 8) were also compared to the LAI retrievals from the regression tree approach discussed in this paper.

4 Results and Analysis

To create the regression tree, homogeneous MODIS LAI samples with the highest quality flag (SCFQC = 0; main algorithm, no saturation) were extracted from the MODIS LAI maps for the

six clear Landsat acquisition dates as listed in Table 1. The thresholds of CV for pure samples were defined as 15% for May 15, June 7, and September 2, and 20% for the rest of inputs to ensure enough samples from each were included. This resulted in 31,929 pure high-quality MODIS LAI samples distributed across the whole Landsat scene area. Of the total set of samples, 50% were randomly selected and used to train the combined regression tree. The average error for the training samples and the remaining 50% of the evaluation samples were 0.15 and 0.16, respectively. The correlation coefficient was 0.96 for both cases. The combined regression tree was applied to all Landsat data, including cloudy scenes. The results were compared to the MODIS LAI products and validated with the field measurements. Temporal trends of LAI from Landsat retrievals were analyzed and compared to the ground observations.

4.1 *Comparison to MODIS LAI*

Landsat LAI maps retrieved from the reference-based approach along with the MODIS LAI from high-quality retrievals (main algorithm, no saturation) for the same period are illustrated in Fig. 2. Black pixels in the MODIS LAI maps represent missing or low-quality retrievals. Although a small percentage of high-quality MODIS retrievals were chosen as training samples (limited by within pixel variation), the MODIS and Landsat LAI are consistent both temporally and spatially, but with greater spatial detail in the Landsat retrievals. While the images from July 8, July 17, and August 26 were not used in sampling and regression tree training, the Landsat LAI retrievals still look similar to MODIS LAI for the cloud-free area. This capability is important for the reference-based approach when there is none or a very limited number of high-quality samples available for a certain date.

In order to compare to MODIS LAI at the pixel level, Landsat LAI was first aggregated to MODIS resolution. Comparisons of Landsat LAI to MODIS were divided into three categories: 1. high-quality MODIS retrievals used in the training process; 2. high-quality MODIS retrievals that have not been used in training due to the high heterogeneity of Landsat surface reflectances within the MODIS pixel; and 3. low-quality MODIS retrievals including retrievals from the backup algorithm (empirical approach) and retrievals under saturation conditions. Note that while the regression tree training process builds a function from the aggregated reflectance [as shown in Eq. (3)], the comparison in Table 2 is based on the aggregated LAI [as shown in Eq. (2)]. Although only homogeneous pixels were considered in the training, there are always small sub-pixel variations of Landsat surface reflectances within MODIS pixels. Therefore, the comparison for training samples (first category) is not exactly the same as a traditional examination of training samples. In this study, only small percentages of homogenous MODIS pixels (5% to 25%) were selected as training samples. The mean bias errors (MBE) between Landsat retrievals and MODIS LAI (Landsat minus MODIS) are between −0.12 and 0.0, suggesting no obvious bias in the combined regression tree. The mean absolute difference (MAD) range from 0.07 to 0.80, which may be partially due to the geolocation errors and mismatch between MODIS and Landsat pixel coverage.

In Table 2, the majority of MODIS retrievals were high-quality retrievals but were not selected as training samples due to high sub-pixel variation in the Landsat surface reflectances. Because they were not used to build the tree, these high-quality retrievals serve as independent references that can be used to validate the combined regression tree. The MBE for these samples ranged from −0.09 to 0.25 for all dates. For the dates that were used for the training, the MBE ranges from −0.09 to 0.12, which is similar to the training samples. Three cloudy scenes that were excluded from training show higher MBE in the range of 0.10 to 0.25. The MAD varies from 0.23 to 0.83, slightly higher than the training samples but still close to the range of MODIS LAI retrieval accuracy (i.e., 0.3 to 0.5). The higher MAD on August 2 is partially due to the high LAI values on that day, as shown in Fig. 2.

The MODIS LAI product also includes retrievals from an empirical approach, which is implemented when the main algorithm fails due to various reasons.[3] These low-quality retrievals from the backup algorithm and the saturated retrievals from the main algorithm comprise the third comparison group. Both MBE and MAD in this group are much larger than the high-quality groups (see Table 2). As data quality in this group is hard to quantify, especially for the saturated retrievals with a LAI higher than 6, they are included in this table only for qualitative comparison purposes.

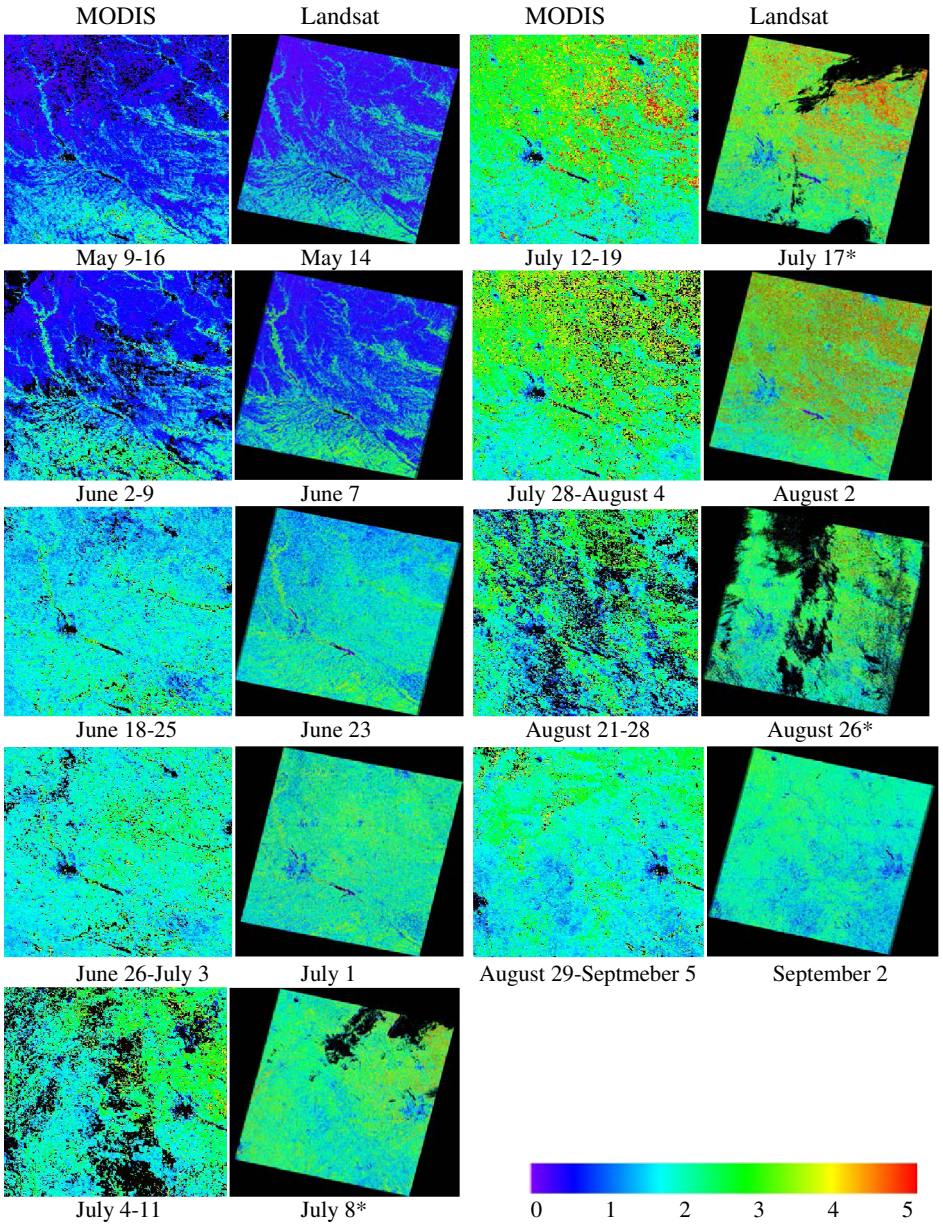


Fig. 2 MODIS LAI product (MOD15A2, high-quality only) and the retrieved LAI from Landsat using MODIS homogeneous LAI samples. Black in the MODIS images represents no data or low-quality retrievals. Black in the Landsat images (*) represents cloudy pixels as detected in the LEDAPS.

The scatter plots between MODIS LAI and Landsat retrievals for three categories are shown in Fig. 3. The darker color represents higher occurrence of LAI values from MODIS and Landsat. Landsat retrievals have been aggregated to MODIS spatial resolution for comparison. Retrievals from all nine Landsat dates were included. Consistent with the statistics in Table 2, the MODIS high-quality LAI products show better agreement with Landsat retrievals and are close to the 1-to-1 line. The MODIS low-quality retrievals show higher variations in scatter plot [Fig. 3(c)].

4.2 *Comparison to Field Measurements*

Estimates of LAI retrieved from this approach were evaluated using SMEX02 ground observations at both the observation scale (10 m) and field scale (>500 m). In Fig. 4, scatter plots are shown between Landsat retrievals and site measurements at the observation scale, which have

Table 2 Mean bias error (MBE; Landsat-MODIS) and MAD between Landsat retrievals and MODIS LAI products, grouped by MODIS retrieval quality. (Portion indicates percentage of pixels in the corresponding category in Landsat image; dashed lines represent cloudy scenes that were not used in training).

Groups	High-quality MODIS LAI (Training samples)			High-quality MODIS LAI (Not training samples)			Saturated or low-quality MODIS LAI retrievals		
Dates	Portion(%)	MBE	MAD	Portion(%)	MBE	MAD	Portion(%)	MBE	MAD
14 May	9.6	0.00	0.07	90.3	0.12	0.23	0.2	−1.91	2.04
7 Jun	16.9	−0.04	0.13	81.7	0.06	0.39	1.4	−0.53	1.21
23 Jun	24.9	−0.06	0.32	74.3	0.07	0.41	0.9	0.27	0.83
1 Jul	5.0	−0.12	0.50	91.8	0.07	0.52	3.2	0.22	0.83
8 Jul	—	—	—	92.7	0.20	0.46	7.3	−1.30	1.65
17 Jul	—	—	—	92.5	0.10	0.74	7.5	0.25	1.01
2 Aug	10.0	−0.04	0.80	78.2	−0.09	0.83	11.8	0.07	1.00
26 Aug	—	—	—	88.5	0.25	0.78	11.5	0.42	1.01
2 Sep	22.2	0.00	0.39	76.5	−0.04	0.39	1.4	−0.31	0.68

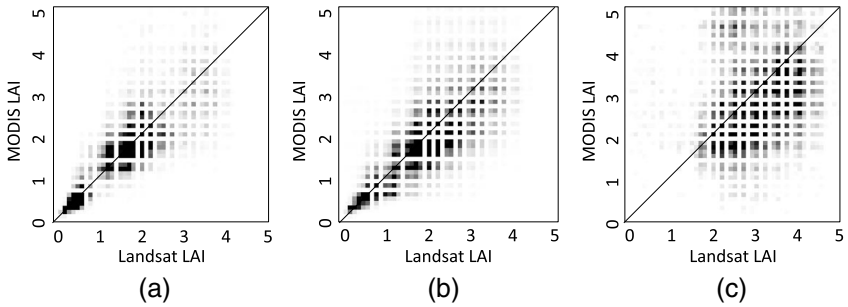


Fig. 3 Scatter plots between MODIS LAI and Landsat retrievals (aggregated to MODIS resolution) for three categories: (a) high-quality MODIS retrievals used in the training process; (b) high-quality MODIS retrievals that have not been used in training; and (c) low-quality MODIS retrievals.

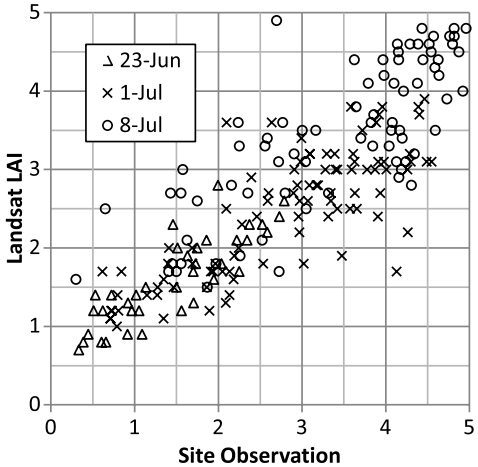


Fig. 4 Scatter plot between Landsat retrievals (30 m) and field observations at the observation scale (10 m) showing good agreement with low to moderate LAI (0–3) but underestimation for high LAI (3–5) for the Landsat retrievals.

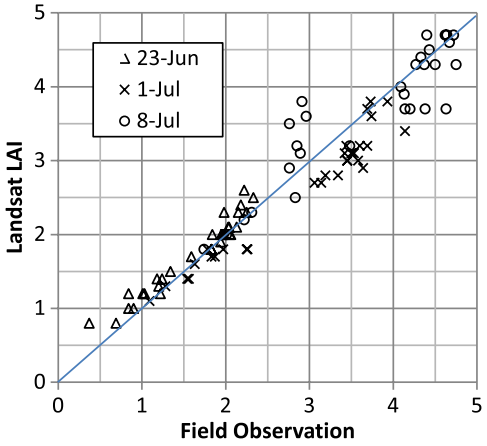


Fig. 5 Scatter plot of LAI between field observations and Landsat retrievals (field observations were extracted from previous publication by Anderson et al.[4]).

been interpolated to the Landsat dates. A better agreement can be seen in the low to moderate range in LAI (0–3). However, under high LAI (3–5), Landsat retrievals underestimate LAI especially for July 1, even though Landsat retrievals and MODIS LAI products agree well in Table 2. The coefficient of determination (r^2) between Landsat retrievals and site measurements is 0.73. An overall MBE (Landsat-field observations) and an MAD are −0.18 and 0.58, respectively. The direct comparison between site measurements (10 m) and Landsat retrievals (30 m) requires that fields are relatively homogeneous at the Landsat scale, which may not be appropriate for some heterogeneous fields.

To reduce uncertainties due to geolocation errors and field heterogeneity that may prevail at the observation scale, Landsat retrievals were averaged and compared at the field scale (>500 m). In Fig. 5, the scatter is shown between the aggregated Landsat retrievals and field-scale LAI estimates from Anderson et al.[4] for June 23, July 1, and July 8. The agreement is improved at the field scale, with a coefficient of determination (r^2) between Landsat retrievals and field measurements at field scale of 0.94 and an overall MBE (Landsat-field observations) of −0.07 and an MAD of 0.23. Similar to Fig. 4, Landsat retrievals underestimate LAI for the high LAI range (3 to 5), especially for July 1. For LAI ranging from 0 to 3, the MBE = 0.09 and the MAD = 0.17. For LAI ranging from 3 to 5, the MBE = −0.28 and the MAD = 0.32.

There may be several reasons for the larger discrepancy when LAI > 3. First, the MODIS eight-day composite LAI product from June 26 to July 3 was used as the reference for Landsat on July 1. Consequently, the MODIS LAI product may be representative of conditions several days prior to the actual Landsat overpass date. For the rapid growth period, this temporal discrepancy can lead to significant differences in LAI. Based on values derived by Anderson et al.[4] the time rate of change in LAI for corn ranged from 0.15 to 0.25 LAI per day for the period of June 23 to July 1. Second, high-quality retrievals and pure MODIS pixels were chosen as the training samples. As the typical crop field size may be smaller than a single MODIS pixel, there may be an unequal distribution of samples, especially for the high LAI values. In addition, high LAI values were more likely to be flagged as the saturated retrievals in the MODIS LAI quality layer and thus be excluded from references. In this paper, there are only about 8.4% of samples having LAI values greater than 3.0. Therefore, a lack of high LAI references may have caused the underestimation by the LAI retrieval technique.

4.3 *Temporal Trends*

Using the combined regression tree, LAI was computed for all nine Landsat scenes listed in Table 1. This provided a dense time-series of LAI at high spatial resolution from mid-May to early September. This period covers the green up, maturity, and partial senescence stages for corn and soybeans in central Iowa. The time-series of LAI from Landsat retrievals and at field scale measurements for all 12 full sampling sites are illustrated in Fig. 6. All sites

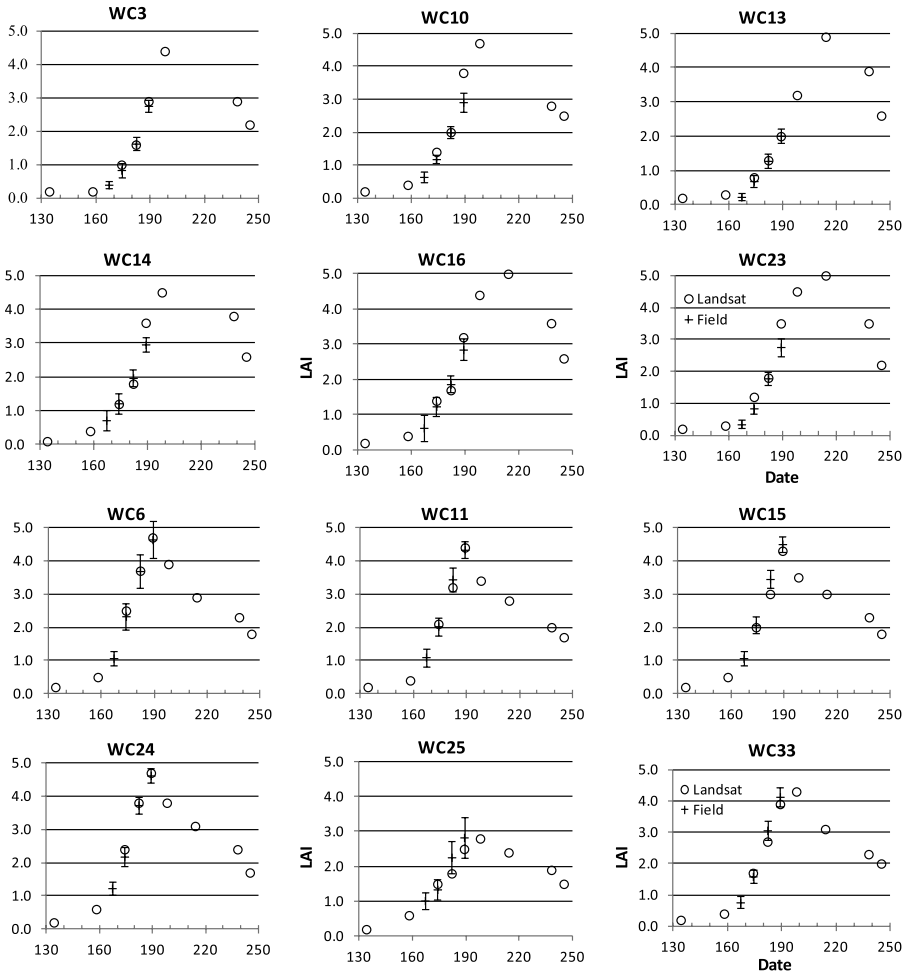
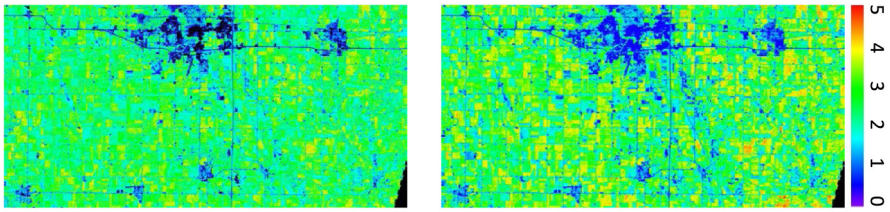


Fig. 6 Time-series LAI from Landsat retrievals (circles) show consistent temporal patterns in comparison with field measurements (crosses) in soybeans (top two rows) and corn (bottom two rows) fields. Error bars show standard deviation of LAI from field measurements.

were sampled in four sampling rounds (crosses in Fig. 6). The LAI retrievals from the Landsat time series show consistent temporal patterns in comparison with field measurements in both corn and soybean field sites and mostly within the error bars (one standard deviation) of field measurements, capturing both green-up time and development trends. The corn sites grew faster and reached the highest LAI on July 8 (DOY 189) image. The soybean sites showed lower LAI during the four SMEX02 measurement rounds but continue increasing up to August 2 (DOY 214). The LAI retrievals for soybean fields show very good agreement over all periods except a few possible overestimates on sites WC10, WC14, and WC23 on July 8. The corn fields show good agreement for the first two rounds (middle and end of June). For the last two rounds, there is some underestimation for corn fields on July 1 and July 8, which can be also observed in Figs. 4 and 5. The depressed growth pattern for site WC25 (corn) due to moisture stress (Anderson et al.[4]) was captured in the Landsat retrievals.

As the Terra and Aqua combined four-day composite MODIS LAI products are only available after day 185 (July 4–7, 2002), the four-day composite products were not used as references in this study, but may result in better agreement with daily field measurements especially during periods of rapid vegetation growth. As an example, Fig. 7 compares results on July 8 for the SMEX02 sites using the MOD15A2 (Terra eight-day composite) and MCD15A3 (Terra + Aqua four-day composite) LAI products as references. Cloudy pixels in the Landsat scene were excluded in the training. Results show higher LAI values are obtained if the four-day composite MODIS LAI is used as the reference compared to the eight-day composite.



(a) Based on MOD15A2.A2002185 (b) Based on MCD15A3.A2002189

Fig. 7 Landsat LAI for July 8, 2002 for the SMEX02 experiment sites as retrieved using (a) Terra eight-day composite MODIS LAI product (185–192, July 4–11) and (b) Terra + Aqua four-day composite LAI (189–192, July 8–11) as reference.

5 Discussion and Conclusions

This paper demonstrates a simple framework for retrieving MODIS-consistent LAI from Landsat reflectance data. The consistent Landsat and MODIS LAI products can be used in the combined ALEXI/DisALEXI approach to produce consistent land-surface flux maps over a range in spatial scales—from continental coverage at 10-km resolution, to local area coverage at 30-m resolution.[31]

The results show that the retrievals from Landsat are consistent with the high-quality MODIS LAI product with no obvious bias. The retrievals were also compared to the SMEX02 field measurements, showing similar temporal patterns. Good agreement between Landsat retrievals and field measurements was observed for LAI in the range of 0 to 3, while higher LAI (3–5) were underestimated especially for corn sites on July 1 and 8.

An empirical regression tree approach was tested and applied in the LAI retrievals. Our experiments show that this approach is flexible and can describe the relationship between surface reflectances and LAI based on intrinsic sample properties. Since this approach does not require an *a priori* relationship between LAI and reflectance, it is important that the quality of the reflectance data is high and that a full range in the reflectance distribution is sampled. If high-quality data are limited, the regression tree approach may not be stable for the LAI retrievals that require significant extrapolation. In such cases, traditional LAI–VI approaches may be considered in the framework. The tests in the current study found no obvious improvement by including additional band combinations (e.g., VIs) in sample training, suggesting that the regression tree approach has the capability of describing complex nonlinear relationships through a series of rules generated from having an extensive set of samples. Increasing the number of rules can improve prediction accuracy but includes the risk of introducing unreliable rules from limited samples. In this paper, we constrained the number of rules to five based on the following observations: 1. the cubist predicted error (or MAD) and correlation coefficient became stable when the number of rules reached 5, and 2. vegetation cover types in our study sites were relatively simple (primarily soybean and corn). We suggest limiting the number of rules to an approximate number of surface types in the study area to avoid over-parameterizing the regression tree procedure.

In this paper, pure pixels were determined by the sub-pixel variations (coefficient of variation) of Landsat surface reflectances within MODIS pixels. This process can be replaced using other homogeneous indexing procedures or high-quality land cover maps if such maps cover the same timeframe as the Landsat observation period. In this paper, a benefit of using the sub-pixel variations is that this approach relies on the reflectance features only rather than predetermined cover types and thus can be performed automatically. A more robust approach to determine homogeneous pixels may improve results and will be explored in future studies.

As the MODIS LAI product used in this paper is an eight-day composite product, the LAI retrievals may not represent the actual Landsat overpass date even though the MODIS production period covers the Landsat acquisition date. The retrieved LAI from Landsat is more like an eight-day composite product rather than a single-day product. This is also revealed in the comparison between Landsat retrievals and MODIS LAI, showing no obvious bias between the two. The example of using a four-day composite (Fig. 7), if available, suggests that a better temporal product can be generated, particularly during the period of rapid crop growth. Also note that

as the MODIS LAI product is used as the reference in this study, this approach is limited to the Landsat scenes acquired in the MODIS-era. A consistent LAI product from AVHRR was proposed recently[5] and may extend the applicability of this approach.

The reference-based approach is a simple empirical technique for retrieving Landsat LAI using MODIS LAI products as a reference. The MODIS LAI samples were retrieved from a radiative transfer model. Therefore, accuracy of this approach depends on MODIS data quality. LAI retrieving accuracy may be lower for small vegetation patches that are not presented at the MODIS pixel resolution. Comparison between direct retrieval using a radiative transfer model and this approach for the same Landsat data was not performed. The application of this approach to more complex landscapes needs to be examined. The combination of LAI-reflectance samples from different dates may introduce additional variations such as vegetation phenology changes and bidirectional reflectance effects caused by different solar geometries. A simple regression tree built from reflectance data may not be able to capture these variations or a more complex pattern of land cover types. Including additional information such as solar geometries, topology, or land cover types in this approach may be necessary for complex landscapes.

By limiting MODIS LAI samples to being both pure and of high quality, LAI information may be lost for certain surface types that appear only at smaller spatial scales (sub-MODIS resolution) and always mixed with other surface types. This may affect the results when MODIS LAI is used as the sole reference in retrieving Landsat LAI, especially when low or high LAI values are missing from the MODIS samples and the regression models need to extrapolate LAI outside the training data range. In this study, high LAI values that were not represented in the range of MODIS samples were encountered in agricultural fields for some Landsat acquisition dates. LAI measurements acquired on the ground within these fields may help to fill this gap, providing LAI information over a full data range of LAI expected at the Landsat scale. The framework in this paper can be modified to combine LAI samples from different data sources. In the cases when MODIS samples cannot represent features from fine-resolution Landsat data, additional information from field measurements may be added. The combination of the two reference sets (MODIS and field measurements) may improve performance in the retrieval over the full range of expected LAI. This option will be explored in future investigations.

Acknowledgments

This work was supported by the U.S. Geological Survey (USGS) Landsat Data Continuity Mission (LDCM) Science Team program and the NASA Earth Observing System (EOS) program. Special thanks go to A. Stern and B. Akhmedou for providing additional LAI field measurements in Walnut Creek watershed published by Dr. P. C. Doraiswamy, and to Dr. P. Beeson for valuable discussions. The U.S. Department of Agriculture (USDA) prohibits discrimination in all its programs and activities on the basis of race, color, national origin, age, disability, and where applicable, sex, marital status, familial status, parental status, religion, sexual orientation, genetic information, political beliefs, reprisal, or because all or part of an individual's income is derived from any public assistance program. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact USDA's TARGET Center at (202) 720-2600 (voice and TDD). To file a complaint of discrimination, write to USDA, Director, Office of Civil Rights, 1400 Independence Avenue, S.W., Washington, D.C. 20250-9410, or call (800) 795-3272 (voice) or (202) 720-6382 (TDD). USDA is an equal opportunity provider and employer.

References

1. P. J. Sellers et al., "Modeling the exchanges of energy, water, and carbon between continents and the atmosphere," *Science* **275**(5299), 502–509 (1997), http://dx.doi.org/10.1126/science.275.5299.502.
2. J. C. Price, "Estimating leaf area index from satellite data," *IEEE Trans. Geosci. Rem. Sens.* **31**(3), 727–734 (1993).

3. R. B. Myneni et al., "Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data," *Rem. Sens. Environ.* **83**(1–2), 214–231 (2002), http://dx.doi.org/10.1016/S0034-4257(02)00074-3.
4. M. C. Anderson et al., "Upscaling ground observations of vegetation water content, canopy height, and leaf area index during SMEX02 using aircraft and Landsat imagery," *Rem. Sens. Environ.* **92**(4), 447–464 (2004), http://dx.doi.org/10.1016/j.rse.2004.03.019.
5. S. Ganguly et al., "Generating vegetation leaf area index earth system data records from multiple sensors. Part 1: theory," *Rem. Sens. Environ.* **112**(12), 4333–4343 (2008), http://dx.doi.org/10.1016/j.rse.2008.07.014.
6. C. O. Justice et al., "An overview of MODIS land processing and products status," *Rem. Sens. Environ.* **83**(1–2), 3–15 (2002), http://dx.doi.org/10.1016/S0034-4257(02)00084-6.
7. Y. Wang et al., "Evaluation of the MODIS LAI algorithm at a coniferous forest site in Finland," *Rem. Sens. Environ.* **91**(1), 114–127 (2004), http://dx.doi.org/10.1016/j.rse.2004.02.007.
8. B. Tan et al., "Validation of Moderate Resolution Imaging Spectroradiometer leaf area index product in croplands of Alpilles, France," *J. Geophys. Res.* **110**, D01107 (2005), http://dx.doi.org/10.1029/2004JD004860.
9. N. V. Shabanov et al., "Analysis and optimization of the MODIS leaf area index algorithm retrievals over broadleaf forests," *IEEE Trans. Geosci. Rem. Sens.* **43**(8), 1855–1865 (2005), http://dx.doi.org/10.1109/TGRS.2005.852477.
10. W. Yang et al., "MODIS leaf area index products: from validation to algorithm improvement," *IEEE Trans. Geosci. Rem. Sens.* **44**(7), 1885–1898 (2006), http://dx.doi.org/10.1109/TGRS.2006.871215.
11. W. Yang et al., "Analysis of leaf area index products from combination of MODIS Terra and Aqua data," *Rem. Sens. Environ.* **104**(3), 297–312 (2006), http://dx.doi.org/10.1016/j.rse.2006.04.016.
12. W. Yang et al., "Analysis of leaf area index and fraction of PAR absorbed by vegetation products from the Terra MODIS sensor: 2000–2005," *IEEE Trans. Geosci. Rem. Sens.* **44**(7), 1829–1842 (2006), http://dx.doi.org/10.1109/TGRS.2006.871214.
13. S. Garrigues et al., "Validation and intercomparison of global Leaf Area Index products derived from remote sensing data," *J. Geophys. Res.* **113**, G02028 (2008), http://dx.doi.org/10.1029/2007JG000635.
14. M. C. Anderson, W. P. Kustas, and J. M. Norman, "Upscaling flux observations from local to continental scales using thermal remote sensing," *Agron. J.* **99**, 240–254 (2007), http://dx.doi.org/10.2134/agronj2005.0096S.
15. M. C. Anderson et al., "Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery," *Hydrol. Earth Syst. Sci.* **15**(1), 223–239 (2011), http://dx.doi.org/10.5194/hess-15-223-2011.
16. M. C. Anderson et al., "A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing," *Rem. Sens. Environ.* **60**(2), 195–216 (1997), http://dx.doi.org/10.1016/S0034-4257(96)00215-5.
17. J. M. Norman et al., "Remote sensing of surface energy fluxes at 10-m pixel resolutions," *Water Resour. Res.* **39**(8), 1221–1237 (2003), http://dx.doi.org/10.1029/2002WR001775.
18. F. Gao et al., "On the blending of the Landsat and MODIS surface reflectance: predict daily Landsat surface reflectance," *IEEE Trans. Geosci. Rem. Sens.* **44**(8), 2207–2218 (2006), http://dx.doi.org/10.1109/TGRS.2006.872081.
19. W. P. Kustas and J. M. Norman, "Evaluating the effects of sub-pixel heterogeneity on pixel average fluxes," *Rem. Sens. Environ.* **74**(3), 327–342 (2000), http://dx.doi.org/10.1016/S0034-4257(99)00081-4.
20. F. Gao et al., "Building consistent medium resolution satellite data set using moderate resolution imaging spectroradiometer products as reference," *J. App. Rem. Sens.* **4**, 043526 (2010), http://dx.doi.org/10.1117/1.3430002.
21. J. G. Masek et al., "A Landsat surface reflectance data set for North America, 1990–2000," *IEEE Geosci. Rem. Sens. Lett.* **3**(1), 68–72 (2006), http://dx.doi.org/10.1109/LGRS.2005.857030.

22. A. Gitelson et al., "Remote estimation of leaf area index and green leaf biomass in maize canopies," *Geophys. Res. Lett.* **30**(5), 1248 (2003), http://dx.doi.org/10.1029/2002GL016450.
23. A. A. Gitelson, Y. J. Kaufman, and M. N. Merzlyak, "Use of a green channel in remote sensing of global vegetation from EOS-MODIS," *Rem. Sens. Environ.* **58**(3), 289–298 (1996), http://dx.doi.org/10.1016/S0034-4257(96)00072-7.
24. A. A. Vina et al., "Comparison of different vegetation indices for the remote assessment of grenn leaf area index of crops," *Rem. Sens. Environ.* **115**(12), 3468–3478 (2011), http://dx.doi.org/10.1016/j.rse.2011.08.010.
25. M. A. Friedl et al., "Scaling and uncertainty in the relationship between the NDVI and land surface biophysical variables: an analysis using scene simulation model and data from FIFE," *Rem. Sens. Environ.* **54**(3), 233–246 (1995), http://dx.doi.org/10.1016/0034-4257(95)00156-5.
26. S. Garrigues et al., "Influence of landscape spatial heterogeneity on the non-linear estimation of leaf area index from moderate spatial resolution remote sensing data," *Rem. Sens. Environ.* **105**(4), 286–298 (2006), http://dx.doi.org/10.1016/j.rse.2006.07.013.
27. P. C. Doraiswamy et al., "Crop condition and yield simulations using Landsat and MODIS," *Rem. Sens. Environ.* **92**(4), 548–559 (2004), http://dx.doi.org/10.1016/j.rse.2004.05.017.
28. J. R. Quinlan, "Learning with continuous classes," in *Proceedings AI'92*, pp. 343–348, World Scientific, Singapore (1992), http://www.rulequest.com/Personal/q.ai92.ps.
29. C. Homer et al., "Development of a 2001 National Land-Cover Database for the United States," *Photogramm. Eng. Rem. Sens.* **70**(7), 829–840 (2004).
30. M. C. Hansen et al., "Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data," *Rem. Sens. Environ.* **83**(1–2), 303–319 (2002), http://dx.doi.org/10.1016/S0034-4257(02)00079-2.
31. M. C. Anderson et al., "Mapping daily evapotranspiration at Landsat spatial scales during the BEAREX'08 Field Campaign," *Adv. Water Res.* (2012), http://dx.doi.org/10.1016/j.advwatres.2012.06.005.

Biographies and photographs of the authors are not available.